REMARKS

The present application was filed on April 27, 2001 with claims 1-24. In the outstanding Office Action dated June 25, 2004, the Examiner has: (i) rejected claims 1-3, 6, 10-12, 18, 20, 21 and 23 under 35 U.S.C. §102(b) as being anticipated by U.S. Patent No. 5,880,788 to Bregler (hereinafter "Bregler"); (ii) rejected claims 4 and 5 under 35 U.S.C. §103(a) as being unpatentable over Bregler, in view of U.S. Patent No. 6,256,046 to Waters et al. (hereinafter "Waters"); (iii) rejected claims 7 and 8 under §103(a) as being unpatentable over Bregler, in view of U.S. Patent No. 6,580,437 to Liou et al. (hereinafter "Liou"); (iv) rejected claims 9 and 22 under §103(a) as being unpatentable over Bregler in view of U.S. Patent No. 6,250,928 to Poggio et al. (hereinafter "Poggio"), and further in view of Liou; (v) rejected claims 13-15, 17, 19 and 24 under §103(a) as being unpatentable over Bregler, in view of U.S. Patent No. 5,884,267 to Goldenthal et al. (hereinafter "Goldenthal"); and (vi) rejected claim 16 under §103(a) as being unpatentable over Bregler in view of Goldenthal, and further in view of Waters.

In this response, claims 2 and 11 have been amended merely to correct minor errors of a typographical nature. Applicants traverse the §102(b) and §103(a) rejections for at least the reasons set forth below. Applicants respectfully request reconsideration of the present application in view of the following remarks.

Claims 1-3, 6, 10-12, 18, 20, 21 and 23 stand rejected under §102(b) as being anticipated by Bregler. The Examiner contends that Bregler discloses all of the elements set forth in the subject claims. Applicants respectfully disagree with this contention. Bregler is directed to the modification of frames of a prerecorded video to create a new video stream which matches essentially any arbitrary utterance (Bregler; column 2, lines 29-32). This methodology requires the "creation of a database of sound-indexed and annotated images" (Bregler; column 4, lines 32-33). The technique disclosed in Bregler "permits any given sound utterance to be substituted for the soundtrack of a previously recorded video sequence, without requiring a video recording of the new sounds being uttered" (Bregler; column 2, lines 19-22; emphasis added). The objectives and methodologies of the present invention set forth in the subject claims are not reasonably analogous to the techniques taught by Bregler.

Independent claims 1, 10 and 18, which are of similar scope, require capturing images of body movements corresponding to one or more words in an utterance and presenting each image segment with a corresponding decoded text word in that utterance. In this manner, the present invention "provides multiple sources of information for comprehending the utterance and allows a hearing-impaired person to quickly and easily ascertain the relationship between body movements ... used to represent the utterance and the corresponding decoded speech" from an automatic speech recognition (ASR) system (Specification; page 2, lines 9-15).

It is to be emphasized that, unlike Bregler, the claimed invention does not utilize images from a previously recorded video sequence and attempt to match arbitrary utterances with the prerecorded images. Rather, the captured image segments used by the claimed invention may be considered analogous to the spectral feature vector set generated by the ASR engine, which are then sent to an image player for presenting each image segment with the corresponding decoded word, as recited in claim 1 (Specification; page 10, lines 23-26). Although Bregler may be capable of synchronizing an arbitrary utterance with a prerecorded image sequence of lip movements stored in a database (Bregler; column 2, lines 35-36), since Bregler uses prerecorded images that are not generated from the actual utterance being presented with the images, Bregler fails to provide multiple sources of information for comprehending and/or verifying the accuracy of the decoded speech, as provided by the present invention.

Specifically, claim 1 requires a visual detector for "capturing images of body movements corresponding to one or more words in the utterance" (emphasis added). Bregler fails to teach or suggest at least this element of claim 1. In this regard, the Examiner contends that Bregler discloses this element at column 4, lines 1-31 and column 7, lines 1-11 (Office Action; page 2, paragraph 4). Applicants respectfully disagree with this contention. While Bregler may disclose, with reference to FIG. 1, analyzing a stored video recording of a person speaking at step S1 "to associate characteristic sounds in the utterance with specific video image sequences," and locating, at step S2, "salient features" in the image sequence (Bregler; column 4, lines 1-6), Bregler fails to disclose capturing images relating to the actual utterance to be presented, as required by claims 1, 10 and 18. In fact, Bregler specifically states (see, e.g., Bregler; column 4, lines 24-25) that a video stream is produced of a person speaking a new utterance (i.e., other than the original utterance from which

the prerecorded video was extracted) by synchronizing a stock video recording to a new soundtrack (Bregler; column 3, lines 66-67). As previously stated, Bregler merely discloses matching an arbitrary utterance to be presented with prerecorded images, "without requiring a video recording of the new sounds being uttered" (Bregler; column 2, lines 19-22). Thus, Bregler fails to disclose capturing images corresponding to one or more words in the utterance being presented, as required by the subject claims.

Likewise, Bregler fails to disclose a visual feature extractor configured for "receiving time information from an automatic speech recognition (ASR) system and operatively processing the captured images into one or more image segments based on the time information relating to one or more words, decoded by the ASR system, in the utterance," as recited in claim 1. In this regard, the Examiner contends that such a feature of the claimed invention is disclosed in Bregler at column 4, line 32 to column 5, line 48 (Office Action; page 2, last paragraph). Applicants respectfully disagree with this contention and submit that while Bregler may disclose image analysis in the general sense at step S2 based on prerecorded video image sequences, Bregler clearly fails to disclose extracting features from the captured image sequence based on the actual utterance to be presented, as recited in claim 1. Because Bregler does not utilize captured images from the actual utterance to be presented, Bregler is simply not capable of achieving an important objective of the claimed invention, namely, comparing and verifying the decoded speech text with the images obtained therefrom in order to determine the accuracy of the recognized text (see, e.g., Specification; page 4, lines 11-17).

Bregler also fails to disclose "an image player operatively coupled to the visual feature extractor, the image player receiving and presenting each image segment with the corresponding decoded word" as recited in the subject claims, wherein the decoded word refers to decoded speech text from the ASR system. Interestingly, the Examiner acknowledges that this feature is absent in Bregler, where he states, in connection with the rejection of claims 7 and 8, that "Bregler fails to specifically disclose that the image segments are displayed with corresponding decoded speech text"(Office Action; page 8, paragraph 3; emphasis added).

For at least the above reasons, Applicants assert that claims 1, 10 and 18 are patentable over the prior art. Accordingly, favorable reconsideration and allowance of these claims are respectfully solicited.

With regard to claims 2, 3, and 6, which depend from claim 1, claims 11 and 12, which depend from claim 10, and claims 20, 21 and 23, which depend from claim 18, Applicants submit that these claims are also patentable over the prior art of record by virtue of their dependency from their respective base claims, which are believed to be patentable for at least the reasons set forth above. Moreover, one or more of these claims define additional patentable subject matter in their own right. For example, claims 2 and 11 further define the image player as being configurable for repeatedly presenting one or more image segments with the corresponding decoded word. In this context, the term "repeatedly" refers to "looping on a time sequence of successive images associated with a particular word(s) in the utterance" (Specification; page 14, lines 13-14). The prior art of record fails to teach or suggest at least this additional feature of the claimed invention.

Likewise, claims 3 and 12 further define the apparatus as including a delay controller for "selectively controlling a delay between an image segment and a corresponding decoded word" in the utterance. Although Bregler may disclose the use of "time warping" to align a new soundtrack to a prerecorded image sequence, Bregler defines time warping as dropping one or more frames from the original recording, "so that the remaining frames in a new video sequence 27 correspond to the timing of the new speech soundtrack 20" (Bregler; column 10, lines 25-30). The concept of time warping taught by Bregler not does involve controlling a delay, and is thus distinguishable from the delay controller recited in claims 3 and 12.

For at least the reasons set forth above, claims 2, 3, 6, 11, 12, 20, 21 and 23 are believed to be patentable over the prior art of record, not merely by virtue of their dependency from their respective base claims, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 2, 3, 6, 11, 12, 20, 21 and 23 are respectfully requested.

Claims 4 and 5 stand rejected under §103(a) as being unpatentable over Bregler, in view of Waters. The examiner acknowledges that "Bregler does not disclose a position detector coupled to the visual detector, the position detector comparing the position of the user with a reference position and generating a control signal . . . and a label generator coupled to the position detector" (Office

Action; page 6, last paragraph). However, the Examiner contends that such features are disclosed in Waters, particularly at column 4, lines 20-41 and column 5, lines 28-59 (Office Action; page 7, paragraph 2). While Applicants may respectfully disagree with this contention, Applicants assert that Bregler and Waters are directed to nonanalogous art, Bregler being directed to the synchronization of prerecorded video image sequences with speech soundtracks (Bregler; column 1, lines 6-7) and Waters being directed to computer vision techniques for sensing a user of a computer system and for providing an interface therewith (Waters; column 1, lines 6-8 and column 2, lines 25-29). Consequently, the Bregler and Waters references, relating to entirely different fields of endeavor, are not believed to be combinable.

Applicants further submit that claims 4 and 5, which depend from claim 1, are patentable over the prior art of record by virtue of their dependency from claim 1, which is believed to be patentable for at least the reasons set forth above. Accordingly, favorable reconsideration and allowance of claims 4 and 5 are respectfully solicited.

Claims 7 and 8 stand rejected under §103(a) as being unpatentable over Bregler, in view of Liou. The Examiner acknowledges that "Bregler fails to specifically disclose that the image segments are displayed with corresponding decoded speech text"(Office Action; page 8, paragraph 3). However, the Examiner contends that such a feature is disclosed in Liou. Applicants respectfully disagree with this contention and submit that Bregler and Liou are directed to nonanalogous art, Bregler being directed to the synchronization of prerecorded video image sequences with speech soundtracks (Bregler; column 1, lines 6-7), as previously stated, and Liou being directed to a video organization and indexing system which uses closed-captioned information of the video to enable content-based abstraction and archival of videos (Liou; column 1, lines 8-12). Consequently, the Bregler and Liou references, relating to different fields of endeavor, are not believed to be combinable as proposed by the Examiner.

Applicants submit that claims 7 and 8, which depend from claim 1, are also patentable over the prior art of record by virtue of their dependency from claim 1, which is believed to be patentable for at least the reasons set forth above. Moreover, these claims define additional patentable subject matter in their own right. For example, claim 7 further defines the apparatus as including a display controller configured to selectively control "one or more characteristics of a manner in which the

image segments are displayed with corresponding decoded speech text." The prior art of record fails to teach or suggest at least this additional feature of the claimed invention. In this regard, the Examiner contends that Bregler teaches controlling a manner in which image segments are displayed with the corresponding audio and Liou teaches that the image segments are displayed with corresponding decoded speech text (Office Action; page 8, paragraph 3). Applicants respectfully disagree with this contention.

Specifically, Bregler fails to disclose selectively controlling a manner in which <u>decoded words</u> from an ASR are displayed with captured image segments corresponding thereto, as required by the subject claims. Liou fails to disclose selectively controlling <u>a manner</u> in which closed-captioned text is displayed with video programming, and thus fails to supplement the deficiencies of Bregler. Claims 7 and 8 are therefore believed to be patentable over the prior art of record, not merely by virtue of their dependency from claim 1, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 7 and 8 are respectfully requested.

Claims 9 and 22, which depend from claims 1 and 18, respectively, stand rejected under § 103(a) as being unpatentable over Bregler in view of Poggio, and further in view of Liou. While Applicants assert that Bregler, Poggio and Liou are not analogous art, as the Examiner suggests, and are therefore not believed to be combinable, Applicants submit that Bregler, Poggio and Liou, when considered in combination, fail to disclose all of the elements set forth in claims 9 and 22. Specifically, contrary to the Examiner's contention in this regard, Poggio fails to disclose an image player displaying "each image segment in a separate window on a display in close proximity to the decoded speech text corresponding to the image segment," as required by claims 9 and 22. The Examiner relies on the disclosure in FIG. 8 and at column 10, lines 52-67 of Poggio as support for such teaching. However, no such teaching exists. Rather, Poggio, in FIG. 8, merely illustrates how a new visual utterance is synthesized from respective visemes. The individual visemes are not displayed to a viewer in separate windows, but "are concatenated, or put together and <u>played seamlessly one right after the other</u>" as part of a video sequence in a single display window (Poggio; column 11, lines 1-4; emphasis added). Therefore, claims 9 and 22 are believed to be patentable over the prior art of record, not merely by virtue of their dependency from their respective base

claims, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 9 and 22 are respectfully solicited.

Claims 13-15, 17, 19 and 24 stand rejected under §103(a) as being unpatentable over Bregler, in view of Goldenthal. With regard to independent claims 13 and 24, which are of similar scope, the Examiner contends that Bregler and Goldenthal, when considered in combination, disclose all of the elements set forth in these claims. Applicants respectfully disagree with this contention. Specifically, the cited prior art fails to teach or suggest at least "capturing a plurality of images representing body movements corresponding to the one or more words in the utterance," as required by claims 13 and 24. The Examiner contends that such a step is disclosed in Bregler at column 4, lines 1-31 and at column 7, lines 1-11 (Office Action; page 10, paragraph 4). However, as previously explained, Bregler merely attempts to match prerecorded video images with words in an arbitrary utterance. To do this, Bregler employs a database of stored image clips and corresponding words. The stored video images are not captured from the utterance being presented. Moreover, Goldenthal fails to supplement the deficiencies of Bregler in this regard.

Furthermore, while Goldenthal may disclose that acoustic-phonetic units are formatted as data records including a "starting time 231, an ending time 232, and an identification 233 of the corresponding acoustic-phonetic unit (Goldenthal; column 4, lines 14-18), neither Bregler nor Goldenthal teaches or suggests "aligning the plurality of images into one or more image segments according to the start and stop times received from the ASR system, wherein each image segment corresponds to a decoded word in the utterance," as required by claims 13 and 24. As previously stated, Bregler may disclose image analysis in the general sense at step S2 based on prerecorded video image sequences, but Bregler fails to disclose extracting features from the captured image sequence based on the actual utterance to be presented. In Goldenthal, starting times and ending times associated with each data record are used to translate the acoustic-phonetic units into visemes by a rendering subsystem (Goldenthal; column 4, lines 20-22). However, the rendering subsystem, like Bregler, does not utilize captures images corresponding to the utterance to be presented. Therefore, Goldenthal fails to supplement the deficiencies of Bregler.

For at least the reasons given above, Applicants assert that claims 13 and 24 are patentable over the prior art. Accordingly, favorable reconsideration and allowance of these claims are respectfully requested.

With regard to claims 14, 15 and 17, which depend from claim 13, and claim 19, which depends from claim 18, Applicants submit that these claims are also patentable over the prior art of record by virtue of their dependency from their respective base claims, which are believed to be patentable for at least the reasons set forth above. Moreover, one or more of these claims define additional patentable subject matter in their own right. For example, claim 14 further defines the method as including the step of "selectively controlling a delay between when an image segment is presented and when a decoded word corresponding to the image segment is presented." Likewise, claim 15 further defines the method as including the step of "selectively controlling a manner in which an image segment is presented with a corresponding decoded word." The prior art of record fails to teach or suggest at least these additional features of the claimed invention.

For at least the reasons set forth above, claims 14, 15, 17 and 19 are believed to be patentable over the prior art of record, not merely by virtue of their dependency from their respective base claims, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 14, 15, 17 and 19 are respectfully solicited.

Lastly, claim 16 stands rejected under §103(a) as being unpatentable over Bregler in view of Goldenthal, and further in view of Waters. The Examiner contends that the combination of Bregler, Goldenthal and Waters discloses all of the elements set forth in claim 16. While Applicants respectfully disagree with the Examiner's contention in this regard, Applicants submit that claim 16, which depends from claim 13, is also patentable over the cited prior art by virtue of its dependency from claim 13. Accordingly, favorable reconsideration and allowance of claim16 is respectfully requested.

Application Serial No. 09/844,120

In view of the foregoing, Applicants believe that pending claims 1-24 are in condition for allowance, and respectfully request withdrawal of the §102 and §103 rejections.

Respectfully submitted,

Date: September 27, 2004

Wayne L. Ellenbogen
Attorney for Applicant(s)
Reg. No. 43,602
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-7662